

RECHERCHE COMPARATIVE DU MODÈLE DE COÛT POUR LA NORMALISATION ET LA DÉNORMALISATION DE BD NoSQL



REALISE PAR : HAO ZHANG

ENCADRE PAR : NICOLAS TRAVERS

CONTEXTE

Dans le BigData, les bases NoSQL répondent au problème de stockage et de distribution des données à large échelle pour souvent répondre à des problèmes de performances en centralisé.

De nombreuses solutions existent, et faire le bon choix lors de la définition des besoins d'un Système d'Information devient un choix crucial, car il est difficile et risqué de faire un changement technologique.

TECHNIQUES



MongoDB, Python et PyMongo

OBJECTIF

Le but de ce projet de recherche est de définir un modèle de coût d'évaluation de requêtes sur des solutions NoSQL.

Ainsi, ce modèle de coût (accès réseaux et accès locaux) sera capable d'estimer pour chaque type de requête ce qu'il en coûtera, et ainsi de choisir la solution NoSQL qui minimise ce coût.

L'objectif de ce projet est donc de raffiner les modèles de coût existant dans la littérature pour avoir un modèle plus global, le plus précis possible.

SOURCE CODE



BASES DES DONNÉES DU TEST

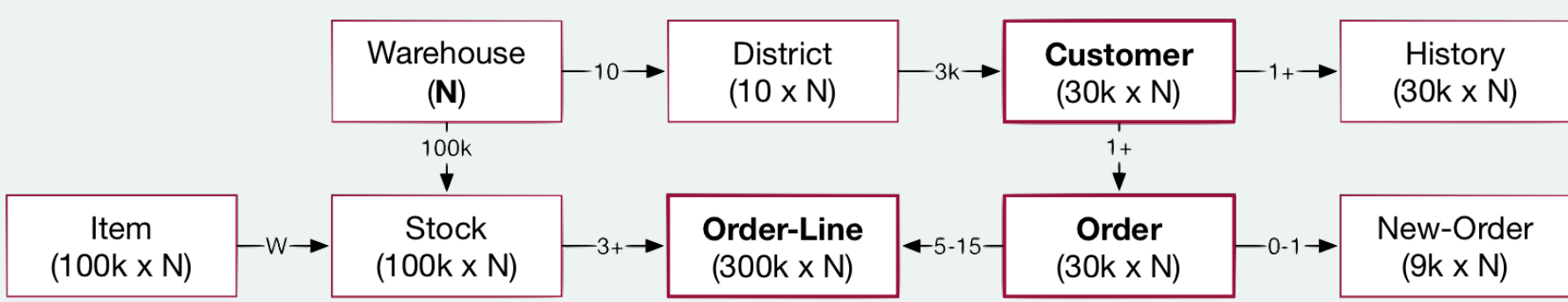
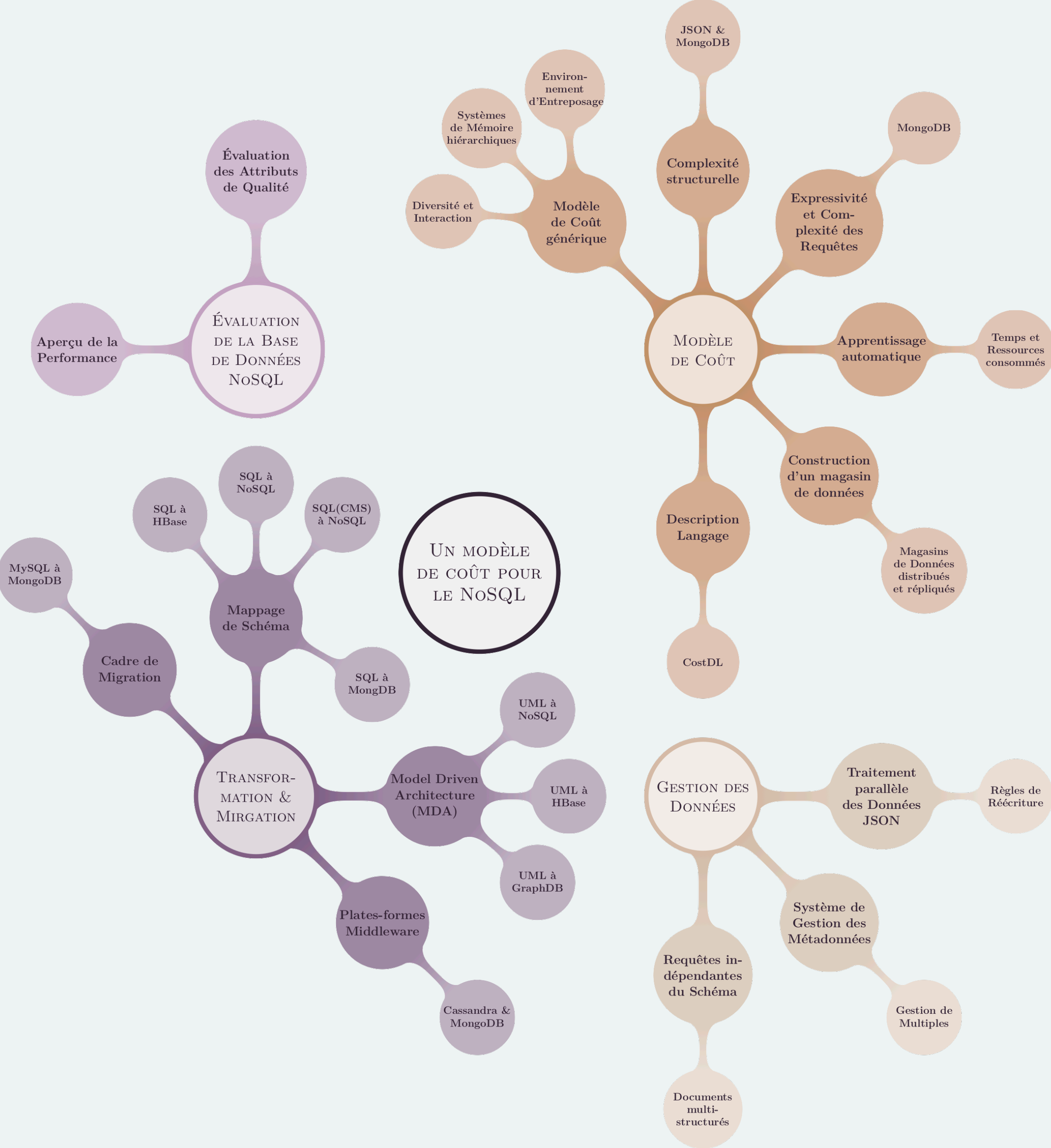


Schéma des collections des bases de données

Semi-Dénormalisation	Dénormalisation
PK o_id	PK c_id
o_d_id	c_d_id
o_w_d	c_w_d
o_c_id	
	customer_order
order_lines	o_id
o_id	o_d_id
o_d_id	o_w_d
o_w_d	o_c_id
o_all_local	order_lines
	o_id
	o_d_id
	o_w_d

Modélisation de semi-dénormalisation et dénormalisation

ÉTAT DE L'ART



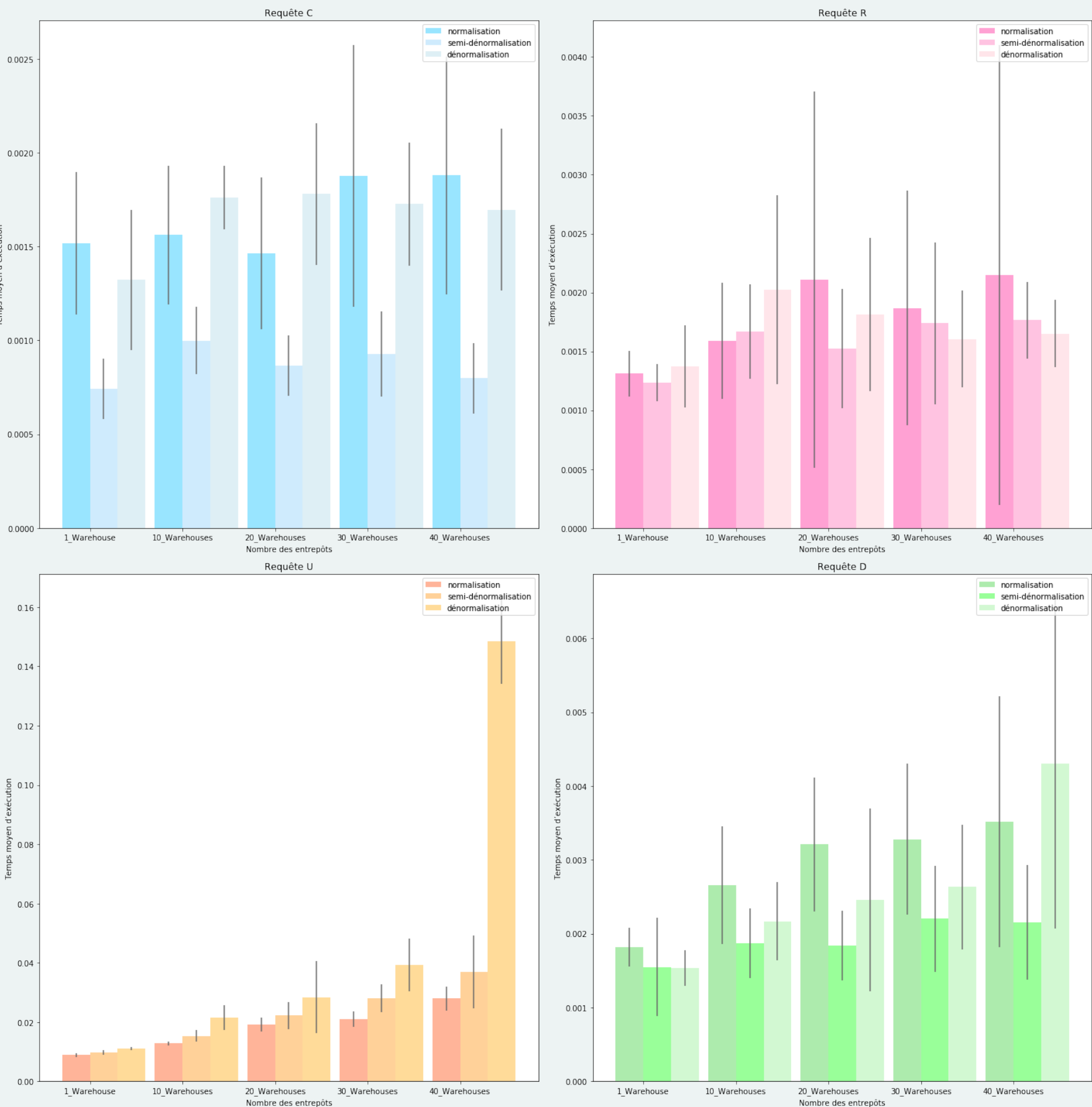
Classification des articles scientifiques

1. Evaluation de la Base de Données NoSQL
2. Transformation et Migration vers NoSQL
3. Modèle de coût d'une base de données
4. Gestion de la base de données NoSQL

REQUÊTES

- C Create : créer une nouvelle commande avec 10 articles différents
- R Read : chercher aléatoirement l'information d'un client par son nom et prénom en montrant l'id de la plus nouvelle commande
- U Update : modifier le temps de livraison des dix articles les plus anciens qui ne sont pas encore livrés
- D Delete : supprimer aléatoirement d'une commande pas encore livrée

RÉSULTATS



Résultats de temps moyen avec l'écart type



Résultats des tests par chaque fois de test